

REPORT

빅데이터 시스템 기말고사 프로젝트

제출일 : 2019.11.05. ~ 2019.11.26. 08:59



Index

- 01) 주제
- 02) 필요성
- 03) 기대효과
- 04) 기존 연구 / 유사 프로그램
- 05) 구현하고자 하는 기능
- 06) 구현된 방법
- 07) 실험 및 결과
- 08) 고찰
- 09) 참 고 문헌
- 10) 느낀점

이름 : 구명희

학번 : 20154215

분반 : 01분반

1. 주제

제가 선택한 주제는 우리나라 사람들의 독서 실태입니다.

통계청의 2017년 데이터 통계에 따르면 우리나라 사람들이 일년에 책을 읽는 빈도는 한달에 한번도 되지 않는다고 합니다. 독서라는 것은 일반독서를 종이책 또는 전자책으로 읽는 것을 말하며 일반독서란 웹소설이나 장르소설까지 포함하는데도 불구하고 우리나라 독서율은 바닥을 치고 있습니다. 저도 사실 군대에서나 책을 읽었지 사회에서는 독서를 즐기는 편이 아니라 독서에 대한 관심을 가져보고자 하는 취지로 이 주제에 대해 조사하게 되었고, 저는 이것의 원인을 규명하고 문제를 해결할 수 있는 방안을 생각해보려고 합니다.

2. 필요성

책보다 훨씬 재미있고 매력적인 멀티미디어가 넘실대는 시대에 책을 읽는다는 것은 절대로 쉬운 일이 아닙니다. 독서는 마음의 양식이라는 말이 있을만큼 많은 사람들이 독서는 중요하다고 말하지만 책을 읽다가 중간에 다른일을 하다보면 책에 대한 생각을 잊어버리게 되는 일도 많을거라고 생각합니다. 저 역시 책을 즐겨 읽지 않아서 우리나라 독서율의 평균을 깎아먹고 있었는데, 기말 프로젝트를 준비하면서 다시 한번 저의 독서율에 대해 돌아보는 계기가 되었습니다.

사실 책을 읽을 시간이 없다던가, 대체할 매체가 있다던가 이런 핑계를 대며 책을 멀리해왔지만 정말 마음만 먹으면 자투리 시간을 활용해서 책 한권 읽을 시간이 없지는 않을 것입니다. 이 조사 데이터를 보고 자신이 조금이라도 문제가 있다는 생각이 든다면 이번주는 도서관에 가지는 않더라도 집에 있는 책을 한권 선택해서 읽어보는 시간을 가져보았으면 좋겠습니다.

3. 기대효과

현재 자신의 독서량을 분명하게 인식하고, 조금이나마 개선시키려는 생각을 심어줄 수 있습니다.

4. 기존 연구, 유사 프로그램

사실 독서의 역사는 그렇게 오래되지 않았습니다. 인류의 읽기능력을 습득한 것은 불과 수천년 전이며, 근대 교육이 도입된 후에도 실현된 대중적인 글 읽기는 즉 겨우 수백년 정도밖에 되지 않았습니다. 인류가 과거에 사냥을 위해 던지기라는 기술을 발달시켜 왔던 것처럼 그것이 시대를 지나 글 읽기라는 기술을 발달시켜온 것입니다. 책 읽기라는 행위가 왜 중요한지에 대한 연구는 지금도 활발히 진행중입니다. 미국 인지과학자 매리언 올프는 “독서는 뇌가 새로운 능력을 학습해 지능을 확대시켜가는 방법을 명확히 보여준다”며 “글을 곧바로 이해하는 능력은 초기 판독에 드는 시간을 줄여주는 대신 더 깊이 분석할 수 있는 시간을 늘리는 데 기여했다”고 말했습니다.

프랑스 인지심리학자 스타니슬라스 드앵은 “종이 위 점과 선이 눈을 거쳐 인간 의식에 심상으로 떠오르고 의미로 이해되는 과정은 경이 그 자체”라고 말하며, 실제로 인간은 한눈에 단어를 이해한다고 생각하지만 뇌는 글꼴에서 의미를 곧바로 얻지 않는다고 합니다. 문자열을 부분으로 쪼개고, 그것들을 다시 문자·음절·형태소 등의 위계로 재구성하는 작업을 거치는데, 이 과정이 모두 자동으로, 무의식적으로 이뤄지기 때문에 모를 뿐이라고 합니다. 그러면서 읽기는 뇌신경에 길을 내고 닦은 결과물이라고 하며, 실험에 따르면 글을 읽을 줄 아는 성인의 뇌와 문맹 성인의 뇌를 비교했을 때 전자가 좌반구 자원을 훨씬 더 많이 이용하고 언어의 기억 폭도 더 커진다고 합니다. “오늘날 뇌과학은 여러 유형의 정보를 조합하고 통합하는 능력이 언어와 연결됐다고 규정한다”며 “인간이 초월적 사고 능력을 갖게 된 건 그 덕분”이라고 설명했습니다.

독서의 연구는 지금도 활발하게 이루어지고 있는데, 미국 예일대의 한 연구팀에서 하루 30분 이상 책을 읽으면 사망 확률이 20%정도나 감소한다고 밝혔습니다. 또한 미국의 피츠버그대학교 연구팀에서는 8~10세의 글을 잘 읽는 어린이와 그렇지 않은 어린이들을 대상으로 뇌 검사를 진행했는데, 아이들이 책을 읽으면 뇌의 구조를 변화시키고 행동조절과 관련이 있는 비만, 뇌졸중 등의 문제를 해결할 수 있으며, 이는 인지 훈련을 통해 뇌의 생물학적 구조를 변화시킬 수 있다는 가능성을 의미한다고 밝혔습니다. 그 외에도 여러 연구팀에서 아이의 뇌 발달에 도움이 된다, 치매 예방에 유효하다, 스트레스 해소에 도움이 된다, 독서를 하면 행복해진다, 우울증을 예방할 수 있다, 다이어트에 도움이 된다 등 많은 분석결과를 보여주고 있습니다.

5. 구현하고자 하는 기능

- 1) 여가 생활중에서 독서의 비율을 알아봅니다.
- 2) 독서하는 사람들의 비율을 알아봅니다.
- 3) 도서관의 수를 알아보고, 도서관의 위치를 지도에 표시합니다.

>> 6. 구현된 방법 에서 자세히 설명하겠습니다.

6. 구현된 방법

• Project-1. 모듈 설치

1) pandas 모듈을 import 합니다.

통계청에서 다운로드 받은 데이터를 pandas 모듈을 사용해서 데이터 프레임 형식으로 저장할 것입니다.

2) matplotlib 모듈과 platform 모듈을 import 합니다.

데이터 프레임 형식으로 저장한 데이터를 matplotlib 모듈에서 지원하는 그래프를 사용해서 분석하려고 합니다.

그래프를 정상적으로 출력해주기 위해 %matplotlib inline 옵션을 설정하고, 한글 폰트를 적용해주었습니다.

3) folium 모듈과 googlemaps 모듈을 import 합니다.

지도에 도서관의 위치를 표시하기 위해 folium 모듈과 googlemaps 모듈을 import 해주었습니다.

• Project-2. 2018년 여가활동 응답 조사하기

1) pandas 모듈을 사용해서 play.xlsx를 데이터 프레임으로 저장했습니다.

다운로드 받은 데이터는 성별, 연령, 학력, 동거가구원수, 혼인상태별 등을 대상으로 조사한 2018년의 여가활동에 대한 응답을 저장하고 있습니다.

2) fillna() 옵션으로 데이터 파일의 존재하지 않는 Nan값을 제거하고 rename() 옵션으로 데이터 파일의 열 이름을 수정했습니다.

3) 데이터 프레임을 다듬기 위해 일부의 행과 열을 제거했습니다.

4) 읽어온 데이터파일에서 연령만 선택한 다음 데이터 프레임을 예쁘게 다듬기 위해 '구분'을 제거하고, '분포'를 인덱스로 설정했습니다.

5) 수정한 데이터 프레임에서 몇 개의 속성만 선택해서 그래프로 그렸습니다.

6) 읽어온 데이터파일에서 가구소득만 선택한 다음 데이터 프레임을 예쁘게 다듬기 위해 '구분'을 제거하고, '분포'를 인덱스로 설정했습니다.

7) 수정한 데이터 프레임에서 몇 개의 속성만 선택해서 그래프로 그렸습니다.

• Project-3.1. 성인 남녀가 종이책을 읽는 비율 조사하기

1) pandas 모듈을 사용해서 paper_person.xlsx를 데이터 프레임으로 저장했습니다.

다운로드 받은 데이터는 2017년의 성별, 연령별, 학력별, 가구소득별로 종이책을 읽었는지 안 읽었는지 조사한 것을 저장하고 있고, 성인 남녀의 데이터만 그래프로 비교해보려고 합니다.

2) fillna() 옵션을 사용해서 데이터 파일의 존재하지 않는 Nan값을 제거하고 rename() 옵션을 사용해서 데이터 파일의 열 이름을 수정했습니다.

3) 읽어온 데이터파일에서 성별만 선택한 다음 데이터 프레임을 예쁘게 다듬기 위해 '구분'과 '사례수'를 제거하고, '분포'를 인덱스로 설정했습니다.

4) 수정한 데이터 프레임을 그래프로 그렸습니다.

• Project-3.2. 성인 남녀가 종이책 또는 전자책을 읽는 비율조사하기

1) pandas 모듈을 사용해서 paper_e_person.xlsx를 데이터 프레임으로 저장했습니다.

다운로드 받은 데이터는 2017년의 성별, 연령별, 학력별, 가구소득별로 책을 읽었는지 안 읽었는지 조사한 것을 저장하고 있고, 마찬가지로 성인 남녀의 데이터만 그래프로 비교하겠습니다.

2) fillna() 옵션을 사용해서 데이터 파일의 존재하지 않는 Nan값을 제거하고 rename() 옵션을 사용해서 데이터 파일의 열 이름을 수정했습니다.

3) 읽어온 데이터파일에서 성별만 선택한 다음 데이터 프레임을 예쁘게 다듬기 위해 '구분'과 '사례수'를 제거하고, '분포'를 인덱스로 설정했습니다.

4) 수정한 데이터 프레임을 그래프로 그렸습니다.

• Project-4. 인구 1인당 평균 독서 권수 조사하기

1) pandas 모듈을 사용해서 reading_book.xlsx 파일을 데이터 프레임으로 저장했습니다.

이제 여가활동중에서 독서의 비율이 얼마정도나 차지하고 있는지 알았으니 이번에는 인구 1인당 평균 독서 권수를 확인하려고 합니다. 여기서 조사하는 독서는 만화나 웹툰 등을 제외한 순수한 일반도서를 읽는 것을 말합니다.

다운로드 받은 데이터는 2009년부터 2017년까지 2년간격으로 조사한 인구 1인당 평균 독서 권수가 저장되어 있으며, 합계, 잡지, 교양서적, 직업 관련 서적, 생활취미 정보 서적, 기타에 대한 column이 저장되어 있는데 하나하나를 비교하는것도 재미있겠지만 프로젝트 진행에 관련이 없다고 생각해서 저는 이중에서 합계만 비교했습니다.

- 2) fillna() 옵션을 사용해서 데이터 파일의 존재하지 않는 Nan값을 제거했습니다.
- 3) 데이터프레임을 예쁘게 다듬기 위해 필요없는 행들을 제거하고, reset_index() 옵션으로 인덱스를 초기화해서 재정렬시켰습니다.
- 5) rename() 옵션을 사용해서 데이터 파일의 열 이름을 수정했습니다.
- 6) 조사할 그룹을 연령, 연령*성별, 교육정도, 혼인상태, 경제활동상태, 가구소득으로 나누고 인덱스를 재설정한다음 각각의 그래프를 그렸습니다.

• Project-5. 인구 1인당 평균 독서 권수에 대한 결론

- 1) 2006년부터 2008년 데이터와 2018년 데이터를 위에서 정리했던 2018년 여가활동 데이터와 비교했습니다.
pandas 모듈로 두 파일을 각각 데이터 프레임으로 저장하고 지금까지 했던 것처럼 데이터 프레임을 예쁘게 다듬었습니다.
다운로드 받은 데이터는 2006년부터 2008년까지의 여가활동에 대한 응답을 저장하고 있습니다.
- 2) 데이터를 비교할 몇 개의 열만 선택해서 비교했습니다.
독서를 포함해서 비교적 높은 수치를 가지는 노래방, 게임 ,인터넷, 쇼핑 등을 선택하겠습니다.
- 3) 인터넷과 독서의 수치만 그래프를 그려서 비교했습니다.

• Project-6. 도서관 이용률

- 1) pandas 모듈을 사용해서 library.xlsx를 데이터 프레임으로 저장했습니다.
다운로드 받은 데이터는 우리나라 사람들이 공공 도서관을 이용하는 비율을 저장하고 있습니다.
- 2) 데이터 프레임을 예쁘게 만들기 위해 fillna() 옵션을 사용해서 데이터 파일의 존재하지 않는 Nan값을 제거하고 rename() 옵션을 사용해서 데이터 파일의 열 이름을 수정했습니다.

• Project-7. 도서관의 위치 알아보기

- 1) pandas 모듈을 사용해서 lotation.xls 데이터 파일을 읽어왔습니다.
다운로드 받은 데이터는 전국의 도서관 이름과 주소, 위치정보 등이 저장되어 있습니다.
- 2) fillna() 옵션을 사용해서 데이터 파일의 존재하지 않는 Nan값을 제거했습니다.
- 3) 읽어온 도서관 목록중에서 광주광역시 광산구와 동구에 위치하는 도서관 리스트들을 선택했습니다.
- 4) 데이터 프레임에 저장된 도서관들의 위도와 경도를 고유 google key를 이용해서 불러온 googlemap에 표기했습니다.

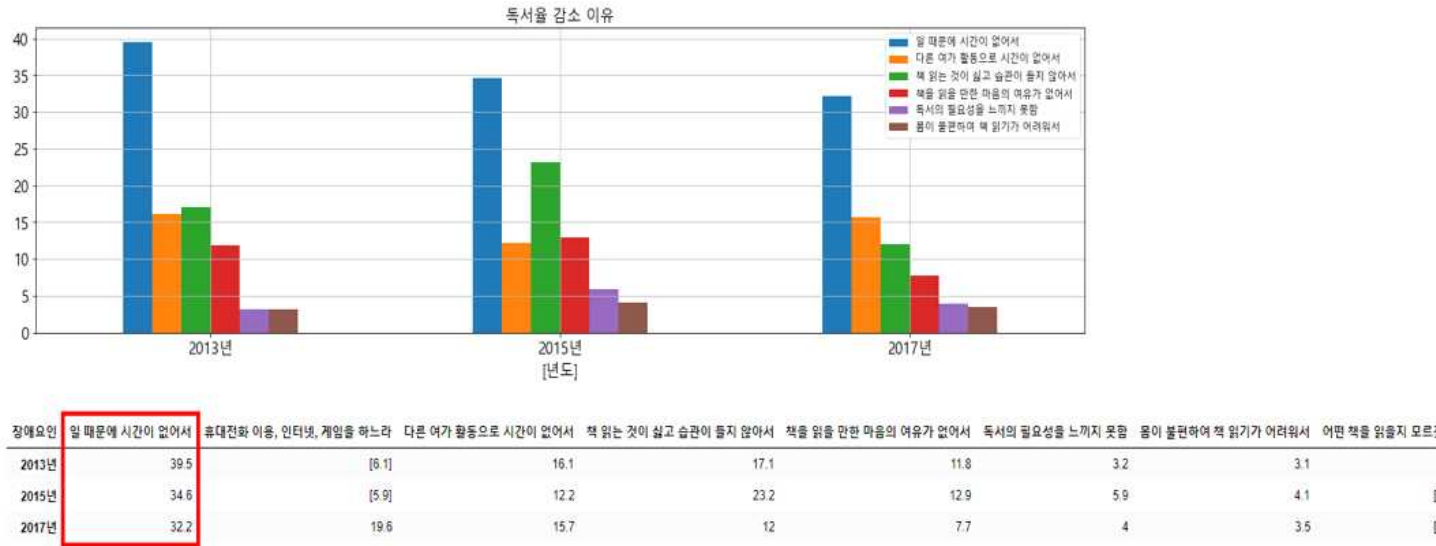
7. 실험 및 결과

비교해본 결과로 2006년, 2007년, 2008년까지의 독서율은 인터넷이 독서 비율보다 낮거나 조금 높았던 것에 비해 2016년과 2018년의 독서율은 인터넷 비율이 독서율의 2.5배를 가볍게 넘기는 놀라운 결과를 보여주고 있습니다.
저는 이 결과를 보고 독서율이 왜 줄어들었는지 의문을 갖고, 독서율 감소 원인에 대한 데이터를 분석해보았습니다.



저는 독서율 감소의 가장 큰 원인이 스마트폰의 보급화라고 생각했습니다. 왜냐하면 2011년~2012년부터 본격적으로 스마트폰이 대중화되기 시작했고, 버스와 지하철에서 스마트폰을 소지하고 다니는 사람들이 많아졌고 스마트폰 중독이라는 단어가 뉴스 기사를 장식하는 빈도가 많아지기 시작한 시점이 바로 2012년이기 때문입니다.

현재 나이에 따른 인구 1인당 평균 독서 권수 그래프의 감소 추세를 보면 2011년까지 독서율이 최고치를 갱신하고 있으며, 그 이후로 독서율이 대체적으로 감소하고 있는 것을 알 수 있습니다. 이를 보고 스마트폰이 가장 큰 원인이라고 생각했는데 의외로 2013년~2017년의 독서율 감소 이유의 조사결과는 ‘일하느라 시간이 없어서’ 항목이 1위였으며, 스마트폰과 관련된 ‘휴대전화, 인터넷게임 때문에’ 항목은 2위를 차지했습니다.



8. 고찰

지금까지 우리나라 독서율의 실태를 조사하면서 알게된점은 크게 4가지입니다.

- 1) 우리의 여가활동 리스트중에서 독서라는 비중이 계속해서 낮아지고 있다는 점.
- 2) 원인은 시간에 치여 책을 읽을 시간이 없고, 책 이외에도 우수한 매체들이 많기 때문이라는 것.
- 3) 그로인해 계속해서 독서율이 낮아지게 될 것이라는 것.
- 4) 독서율을 증가시킬 수 있는 좋은 방법은 도서관 이용률을 증가시키면 된다는 것.

계속해서 낮아지고 있는 독서율은 우리나라에만 해당되는 문제가 아니라 세계적인 이슈라고 합니다. 하지만 아무리 노력해도 종이책의 독서율은 정보화 시대의 발전에 따라 계속해서 낮아질 수밖에 없습니다. 전자책을 보면 되는 것이 아닌가라는 의문을 갖는 사람도 있겠지만, 아직 종이책에 비해 전자책의 출간 수는 턱없이 부족하다고 합니다. 특히 우리나라는 외국시장에 비해 상당히 뒤쳐져 있으며 애초에 책자체를 잘 읽지 않아서 특정 분야의 도서만 출판되고 있는 실정이라고 합니다. 물론 전자책은 종이책에 비해 많은 이점이 있지만 그럼에도 불구하고 아직까지는 전자책이 완벽하게 종이책의 대체제가 될 수 없을 것 같습니다.

독서를 남 주기 위해서 하는것은 아니지만 정부 차원에서 문제가 되는걸 인지하고 있다면 해당 문제에 대해 위기감과 책임감을 갖고 여러 방향으로 독서율을 늘리기 위한 노력을 보여줬으면 좋겠습니다. 예를 하나 들자면, 현재 독서율은 노인층으로 갈수록 줄어들고 있는 실정입니다. 당장 저희 부모님만 봐도 노안으로 스마트폰의 기본 폰트조차 잘 안보이지 않아서 가장 큰 폰트 크기로 바꿔 사용중입니다. 이런 분들이 책을 읽기란 곤란할 것입니다. 어른들을 대상으로 안경점과 제휴?해서 돋보기를 나눠주는 등 다양한 지원을 해주고 독서 이벤트를 진행하는 방안 등이 있었으면 좋겠습니다.

9. 참고 문헌(사이트)

문헌은 참고하지 않았고 사이트 위주로 참고했기 때문에 사이트 링크를 제출하겠습니다.

8. 고찰을 작성하기 위해 01)독서의 중요성을 참고했으며 02)~10) 목록은 다운로드 받았던 통계청 데이터 링크입니다.

01) 독서의 중요성	1) https://news.samsung.com/kr/책-안-읽는-시대-그럼에도-불구하고-독서가-필요 2) https://blog.naver.com/eggshamegg/220783800562
02) 여가활동별 참여비율 (2016~2018)	http://kosis.kr/statHtml/statHtml.do?orgId=113&tblId=DT_113_STBL_1028375&conn_path=I3
03) 여가활동별 참여비율 (2006~2008)	http://kosis.kr/statHtml/statHtml.do?orgId=113&tblId=DT_113_STBL_1014277&conn_path=I3
04) 독서량 (2009~2017)	http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1SSCL020R&conn_path=I3
05) 독서장애요인 변화추이	http://kosis.kr/statHtml/statHtml.do?orgId=113&tblId=DT_113_STBL_1024028&conn_path=I3
06) 성인의 인근지역 공공도서관 이용률/이용빈도	http://kosis.kr/statHtml/statHtml.do?orgId=113&tblId=DT_113_STBL_1024369&conn_path=I3
07) 성인이 공공도서관을 이용하지 않는 이유	http://kosis.kr/statHtml/statHtml.do?orgId=113&tblId=DT_113_STBL_1015728&conn_path=I3
08) 종이책+전자책 독서율	http://kosis.kr/statHtml/statHtml.do?orgId=113&tblId=DT_113_STBL_1025546&conn_path=I3
09) 종이책 독서율	http://kosis.kr/statHtml/statHtml.do?orgId=113&tblId=DT_113_STBL_1025544&conn_path=I3
10) 공공 도서관 위치	https://www.data.go.kr/dataset/15013109/standard.do

10. 느낀점

이번 프로젝트에서는 제가 직접 개인 프로젝트의 주제를 정하고 기능을 어떻게 구현할지 처음에 설계하면서 여러 가지 기능을 구현하려고 생각했었는데, 프로그램을 구현하면서 상상했던것만큼 구현할 수가 없었습니다. 책에서 나온 기능들을 최대한 사용해서 진행하려고 했는데, 시간적으로나 능력적으로나 부족한 부분이 많아 몇가지는 구현을 포기하고 간단한 기능으로 대체한 것이 많아서 아쉬웠습니다.

특히 3장의 웹페이지에 대한 정보를 읽어와서 위치를 지도에 표기하는 것을 해보고 싶었는데, 놀랍게도 공공 데이터 센터에서 전국의 도서관별 위도와 경도를 제시한 데이터가 있어서 간단하게 그 위도와 경도를 사용해서 지도에 위치를 표시하는 것으로 대신했던 점과, 7장의 시계열 데이터를 꼭 사용해보고 싶었는데 데이터가 너무 부족해서 진행할수 없었던 예로사항이 있었습니다.

이번 학기 프로젝트는 제가 python 언어에 대한 이해도가 부족해서 진행에 걱정이 많았었는데, 막상 만들어서 결과물을 보니 개인적으로는 만족스러운 결과가 나온 것 같습니다.

마지막으로, 교수님 한학기동안 고생하셨습니다!